

Prediction Method of Stock Fluctuation Trend Based on LASSO-BPNN

Hongzheng Li^{1,*}

¹*Rutgers Business School, Rutgers University, Newark, New Jersey 07103, China*

^{*}*Corresponding author. Email: hl877@scarletmail.rutgers.edu*

ABSTRACT

As an important part of the securities industry, the stock market has always attracted the attention of investors. In view of the fact that there are many factors affecting the stock market and it is difficult to predict, this paper proposes a prediction method of stock fluctuation trend based on LASSO-BPNN. Firstly, the influencing factors of stock fluctuation trend are analyzed, and the variables which have significant influence on stock price are selected by LASSO method. Secondly, on the basis of variable selection, the prediction model of stock fluctuation trend based on LASSO-BPNN is constructed to predict a new energy stock in China's A-share market. Finally, empirical analysis is carried out to verify the accuracy of the method. The empirical results show that the prediction method proposed in this paper accurate for the stock fluctuation trend judgment, which is suitable for the analysis of stock market. According to the prediction results, some investment suggestions are put forward at the end of this paper.

Keywords: *LASSO-BPNN, prediction method, stock fluctuation trend, China's A-share market, empirical analysis*

1. INTRODUCTION

With the rapid development of economy, more and more investors tend to invest their money in the stock market. Affected by many factors, the stock market has strong volatility, so it has become an investment place with high risk and high return. The goal of stock investment is to reduce the investment risk as much as possible on the premise of stable income, and to improve the return as much as possible under the premise of stable risk, which requires the use of effective analysis methods and means to comprehensively consider the influencing factors of the stock market, and then make an objective and accurate judgment. There are many factors that can affect the stock price, so we can't consider the stock as a simple time series. Therefore, it is very important for investors to find a stable and effective stock investment method, reduce investment risk and improve return efficiency under the premise of considering various factors related to stock price.

The research on stock data started from a very early time, and scholars at home and abroad have been trying to use different methods to analyze it. The traditional time series analysis methods mainly include ARIMA model and conditional heteroscedasticity model.

However, due to the complexity of the stock market, the accuracy of the traditional time series analysis method is difficult to meet the requirements of the existing prediction accuracy. With the development of computer technology, especially artificial intelligence technology, neural network has been developed rapidly. Many scholars have applied it to the research of stock price forecasting. Zhou Qin et al. [3] used the combination method of BP neural network and grey model to predict stock price. In this method, the predicted value of grey model was taken as the input variable of BP neural network, and the actual value of stock price was taken as the output variable of the network. Hu Lingyan et al. [4] used the stepwise regression model and BP neural network to predict the prices of the two stocks, and analyzed and compared the accuracy of the prediction results by using the average relative error and the mean square error, indicating that the two models have a certain effect on the short-term prediction of stock prices. Nemes et al. [5] used data mining technology to process data and designed neural network model to predict the short-term fluctuation of stock price. However, the input variables used in these methods do not comprehensively consider many complex factors such as market fluctuation, policy influence, enterprise finance, etc., and the final prediction results of neural

network have a great correlation with the selection of input variables, among which the selection of input variables is the most difficult. Therefore, this paper introduces LASSO analysis method to analyze the variables that have great influence on the stock trend, and then inputs them into BP neural network as input variables to predict the stock fluctuation trend, and verifies its effect through empirical analysis.

2. LASSO METHOD AND VARIABLE SELECTION

2.1. LASSO Method

With m independent variables x_1, x_2, \dots, x_m and dependent variable y , the linear model is established as follows

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon \quad (1)$$

Where α is a constant term, $\beta_1, \beta_2, \dots, \beta_m$ are regression coefficients, ε is a random perturbation term.

Setting $(x_{i1}, x_{i2}, \dots, x_{im}), i=1, 2, \dots, n$ be the n sets of observations of the variable, it is assumed that the data has been centralized and standardized, so

$$\begin{cases} \sum_{i=1}^n y_i = 0 \\ \sum_{i=1}^n x_{ij} = 0, j=1, 2, \dots, m \\ \sum_{i=1}^n x_{ij}^2 = 0 \end{cases} \quad (2)$$

and $\beta = (\beta_1, \beta_2, \dots, \beta_m)^T$. Then ASSO estimation of unknown parameters α and β in model (1) is defined as

$$(\bar{\alpha}, \bar{\beta}) = \operatorname{argmin} \left[\sum_{i=1}^n (y_i - \alpha - \sum_{j=1}^m \beta_j x_{ij})^2 \right] \quad (3)$$

Where, $\sum_{j=1}^m |\beta_j| \leq s$. Here, $s \geq 0$ is the penalty parameter. Let $\bar{\beta}_j^0$ be the solution of the least squares estimate, $s_0 = \sum_{j=1}^m |\bar{\beta}_j^0|$. When $s \geq s_0$, the optimal solution in equation (2) is the least square solution. When $s < s_0$, some regression coefficients will be reduced or tend to 0 or even equal to 0. These variables equal to 0 will be deleted to achieve the purpose of variable selection. Since the data has been centralized and standardized, for any $s > 0$, the solution of α in equation (3) is $\bar{\alpha} = 0$. Therefore, equation (3) can be equivalent to

$$\bar{\beta} = \operatorname{argmin} \left[\sum_{i=1}^n (y_i - \alpha - \sum_{j=1}^m \beta_j x_{ij})^2 \right] \quad (4)$$

Where, $\sum_{j=1}^m |\beta_j| \leq s$. Each value of s corresponds to a

lasso solution through equation (4). After several steps, all LASSO solutions with different s values can be obtained.

In order to estimate the parameter s effectively, Akaike information criterion (AIC) is used to estimate the parameter s . AIC is a criterion to measure the goodness of statistical models. It is based on the concept of entropy. It can balance the complexity of the estimated model and the goodness of the model fitting data. In general, AIC can be expressed as

$$AIC = 2K - 2\ln(L) \quad (5)$$

Where, K is the number of parameters and L is the likelihood function.

If the error of the model obeys the independent normal distribution, let n be the observation number, and RSS is the sum of the residual squares, then AIC can be expressed as

$$AIC = 2K + n\ln(RSS/n) \quad (6)$$

In the above equation, the number of parameters is increased to improve the goodness of fitting. The smaller the AIC value, the more accurate the model is. Therefore, the preferred model should be the one with the smallest AIC value.

2.2. Data Sources and Variables Selection

The data sample of this paper uses new energy stock, whose code is 601012 in China's A-share market, to study the monthly closing price data of 103 months from May 2012 to November 2020. Nine factors including per capita GDP (F1), oil price (F2), market environment (F3), consumer confidence index (F4), consumer price index (F5), sales price (F6), coal production (F7), consumer attitude (consumer expectation index F8) and after-sales service (consumer satisfaction index F9) are preliminarily given.

Lasso algorithm is used to select the influencing factors of the reconstructed new energy stock, the results are shown in Table 1.

Table 1 LASSO variable selection results

Explained variable (stock price)	Explanatory variables (influencing factors)
X	F3
	F4
	F6
	F7
	F8

We implement the choice of influencing factors based on AIC and get the regression equation of stock

price and influencing factors, and the estimation formula is as follows

$$X = 0.475F_4 - 0.526F_3 - 0.672F_7 + 0.586F_6 - 0.174F_8 \quad (7)$$

According to the experimental results of stock analysis, from formula (7), we can see that the consumer confidence index and sales price are positively related, and coal output, market environment, consumer attitude are negatively related to it, and the influence degree of these factors is: coal production > sales price > market environment > consumer confidence index > consumer attitude. From the above analysis, we can conclude that when investors invest in new energy stocks, they can properly analyze these influencing factors, so as to make more profitable investment decisions.

3. LASSO-BPNN PREDICTION MODEL

3.1. BPNN

Back propagation neural network (BPNN) is a network model trained by error back propagation algorithm and is one of the most widely used neural network models. BPNN can represent extremely complex nonlinear systems. BPNN is composed of three parts: input layer, at least one hidden layer and output layer. Each layer is composed of one or more neural nodes. The neural nodes of each layer only receive the input from the neurons of the previous layer. The input information must be processed by the neurons of each layer before it becomes the output of the output layer. That is, the input layer receives the external input mode, and the nodes of the hidden layer only receive the input layer information. Information input, the nodes in the output layer only accept the information input from the hidden layer. The structure of BPNN is shown in Figure 1.

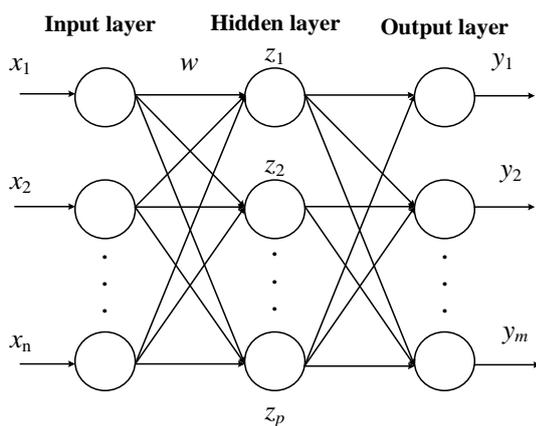


Figure 1 BPNN structure

The main idea of BP algorithm is to divide the learning process into two stages: the first stage is the positive propagation of information, which goes through the input layer, hidden layer and output layer in turn, and finally calculates the actual output value of each neuron. The second stage is error back propagation.

When there is a large error between the actual output value and the expected value, the feedback information propagates in the path of output layer, hidden layer and input layer. Specifically, it is to allocate the errors to the neurons of each layer, so as to obtain the error signals of each layer of neurons, and take these error signals as the basis for correcting the weights of each connection. So that the final error of the two processes is minimized. In fact, when the error reaches the desired goal, the connection weights and neuron threshold of the network reach the best combination, and the learning process of the network is over. BPNN is widely used in prediction field because of its good learning and anti-interference ability.

3.2. Combination Model

In the practice of prediction, the object of prediction is likely to be a more complex system, which is affected by many complicated factors. The stock fluctuation trend studied in this paper is such a complex system. A single forecasting method may not achieve satisfactory results. Based on the above considerations, this paper proposes a combined forecasting method based on LASSO-BPNN to predict the stock fluctuation trend.

In the new energy stock fluctuation trend prediction model, five variables, F3, F4, F6, F7 and F8, are selected as the inputs of BP neural network. The output of the network is the closing price of the stock in the current month. The training sample data of the prediction model still selects the monthly stock closing price from May 2012 to may 2020, and the test sample selects the monthly stock closing price from June 2020 to November 2020.

In the prediction model, the number of input layer neurons is 5, and the number of hidden layer nodes is selected from 3. According to the prediction results, the number of hidden layer nodes is gradually adjusted, and the number of output layer neurons is 1. At the same time, the transfer functions of hidden layer and output layer are set as tansi and purelin, and the training function is set as trainlm. The number of network training is set to 10000, and the target error is set to 0.00001. The structure of the combined prediction model is shown in Figure 2.

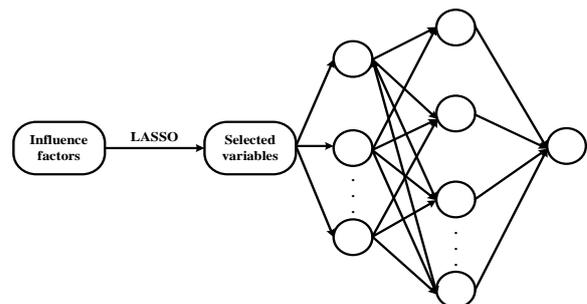


Figure 2 The structure of the combined prediction model

4. EMPIRICAL ANALYSIS

When the number of hidden layer nodes is set to 10, the network training is good. On this basis, the data of

test samples are input into the trained combination forecasting model. Compared with the prediction results of BPNN, the prediction results of the proposed method are shown in Table 2.

Table 2 Prediction results

Time	Closing price	BPNN	Relative error/%	LASSO-BPNN	Relative error/%
2020.06	40.73	35.24	13.48	38.72	4.93
2020.07	56.80	62.36	9.79	61.74	8.70
2020.08	63.28	53.62	15.27	69.96	7.40
2020.09	75.01	81.24	8.31	78.23	4.29
2020.10	75.99	83.56	9.96	80.21	5.55
2020.11	68.7	75.62	10.07	74.36	8.24
Average relative error/%		11.15		6.52	

From the prediction results of BPNN, the prediction error of BPNN is about 10%, and the average prediction error is 11.15%. As for the fluctuation trend of stocks, BPNN accurately forecasts the rising and falling trend of stock prices from June to July, August to September, September to October and October to November in 2020, but misjudges the fluctuation trend of stock prices from July to August. The average prediction error of LASSO-BPNN is about 6.52%, and it can accurately predict the stock fluctuation trend within 6 months, which has higher prediction accuracy and is more suitable for analyzing the stock fluctuation trend.

According to the relative error analysis of new energy stock forecast, the combined forecasting accuracy of lasso method and BPNN is obviously higher than that of single BPNN. Therefore, we can use the combination of lasso method and BPNN to predict the stock price of new energy sector in China's A-share market.

5. CONCLUSION AND INVESTMENT SUGGESTIONS

5.1. Conclusion

Guided by the theories of economics, statistics, econometrics and other disciplines, this paper first makes an overall analysis of the background of the current stock price forecast, then makes an empirical analysis of the influencing factors of stock price by using lasso statistical method, and finally forecasts the fluctuation trend of stock price by combining BP neural

network on the basis of variable selection. The main conclusions of this paper are as follows:

(1) As an emerging industry, although the stock price of new energy is on the rise in general and increases rapidly in some time periods, it may fluctuate violently in the short term due to various factors. Therefore, it shows that the industrial structure is not fully mature, but the future development prospects are good. Therefore, long-term investment can be considered, but blind pursuit should be avoided.

(2) This paper selects nine quantifiable variables from a variety of stock price factors, and uses lasso method to select variables for new energy stocks. The empirical results show that: market environment (F3), consumer confidence index (F4), sales price (F6), coal production (F7), and consumer attitude (consumer expectation index F8) are the main influencing factors of new energy stock price. And the influence decreased in the order of coal production > sales price > market environment > consumer confidence index > consumer attitude.

(3) In the research of new energy stock price prediction, BPNN, LASSO-BPNN combination forecasting method are mainly used. The prediction accuracy of LASSO-BPNN is higher than that of BPNN. Using lasso method and BPNN combination forecasting method to forecast the price of new energy stock can better predict the future trend of new energy stock.

5.2. Investment Suggestion

For the investment of new energy stocks, this paper suggests that investors should hold the relevant stocks of new energy appropriately. As the development trend of new energy in the future, there is a large room for appreciation in the future. However, when investors invest in related stocks, they should consider a variety of factors for a single stock in the new energy sector. They should not blindly chase up and down. They should reasonably adjust their share of stocks, and form a long-term investment system with industry leading stocks as the main and industry emerging stocks as the auxiliary. In addition, due to the variety of new energy, we should reasonably analyze the potential energy types in the industry in combination with the real world economic development, and make rational investment.

REFERENCES

- [1] Y. Chen, W. Lin and J. Z. Wang, A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features, *IEEE Access.* 7, (2019) 148047-148058.
- [2] X. Yuan, J. Yuan, T. Jiang and Q. U. Ain, Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market *IEEE Access.* 8 (2020) 22672-22685.
- [3] Q. Zhou, D. Yong, X. Wang, Application of grey neural network in stock price forecasting. *Business,* 39 (2015) 169-170.
- [4] L. Hu, J. Zhu, Comparative Analysis of Stock Price Forecast Based on Stepwise Regression and BP Neural Network, *Journal of Liaoning Institute of Technology (Natural Science Edition).* 39(3) (2019) 201-205, 210.
- [5] MD. Nemes, A. Butoi, Data Mining on Romanian Stock Market Using Neural Networks for Price Prediction, *Informatica Economica.* 17(3) (2013) 125-126.